2005

Simultaneous Equations in Ordered Discrete Responses with Regressor-Dependent Thresholds

Myoung-jae Lee
Singapore Management University, mjlee@smu.edu.sg

A. Kimhi

DOI: https://doi.org/10.1111/j.1368-423X.2005.00159.x

Econometrics Journal (2005), volume **8**, pp. 176–196.

Simultaneous equations in ordered discrete responses with regressor-dependent thresholds

MYOUNG-JAE LEE*† AND AYAL KIMHI‡

**School of Economics and Social Sciences, Singapore Management University, Singapore*

†Department of Economics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong
E-mail: mjlee@smu.edu.sg

‡Agricultural Economics Department, Hebrew University, Rehovot 76100, Israel
E-mail: kimhi@agri.huji.ac.il

Received: October 2003

Summary The parameters of ordered discrete response (ODR) models are identified only up to a positive scale. In this paper, we examine the identification issue for simultaneous equations with ODR, where the well-known identification problem in simultaneous equations of recovering structural-form parameters from reduced-form parameters is compounded with the ODR identification problem. We allow the thresholds in ODR to be regressor dependent as well as constant; the former is particularly challenging because threshold parameters get mixed with regression parameters, adding one more dimension to the identification problem. We also explore a cross-equation restriction on threshold differences, under which the structural form parameters are fully identified as if the dependent variables are continuously distributed. An empirical example with farm–household joint labour supply is provided to illustrate the identification issues, to show how our proposals work and to apply tests devised for the threshold constancy and cross-equation restrictions.

Keywords: *Identification, Ordered discrete response, Simultaneous equations.*

1. INTRODUCTION

Consider a simple ordered discrete response (ODR) model

$$y_i = \sum_{r=1}^{R-1} 1\left[y_i^* > \gamma_r\right] = \sum_{r=1}^{R-1} 1[x_i'\eta + v_i > \gamma_r], \quad i = 1, \ldots, N, \tag{1}$$

where $1[A] = 1$ if A holds and 0 otherwise, y_i^* is a latent continuous response variable, x_i is a $k \times 1$ vector of explanatory variables, η is a vector of parameters, v_i is an error term and γ_r', $r = 1, \ldots, R-1$, are (unknown) thresholds; y_i takes the values $0, 1, \ldots, R-1$ (R-many categories). For example, y_i may denote an income category whereas y_i^* is the actual income; here, γ_r's are the known thresholds for the income categories. Alternatively, the thresholds γ_r could be unknown, for example, if y_i is the extent of work (none, part-time and full-time) while

y_i^* is the number of hours worked per day. Another example is y_i as a job rank and y_i^* as a latent continuous measure of promotability; here, γ_r's can vary across i if there is discrimination based on individual characteristics such as gender or race. In this case, the notation γ_{ri} would be more appropriate for the thresholds.

Since any non-decreasing transformation can be applied to both sides of $x_i'\eta + v_i > \gamma_r$ in (1) without changing the inequality, η is not fully identified, and this identification problem becomes more complicated if we have varying thresholds γ_{ri} instead of the constant γ_r. Even more complicated is the case of multiple ODR equations that are simultaneously related, because the identification problem of each ODR equation is entangled with the classical identification problem in simultaneous equations of recovering the structural-form (SF) parameters from the reduced-form (RF) parameters.

In this paper, we examine the identification issue in ODR simultaneous equations. We do this first for constant thresholds; a by-product of this is a simpler presentation of order and rank conditions than is typically found in econometrics textbooks. Second, we allow for regressor-dependent (i.e. varying) thresholds, which further complicates the identification issue because the RF threshold parameters get mixed with the RF regression parameters. Third, we show that the identification problems are relieved by taking advantage of a cross-equation restriction on threshold differences, because the threshold differences are informative for the scale of y_i^*, which would not be available if the observed responses were binary.

Kimhi and Lee (1996) estimated simultaneous equations with ODR, but had difficulty interpreting the magnitudes of the SF parameters due to the ODR identification problem. Windmeijer and Santos-Silva (1997) analyzed a count variable (number of visits to doctors) with regressors including a binary health indicator (1 if poor/fair and 0 if good/excellent). In their Poisson-type count-response framework, simultaneity is not allowed, but a count response can be modelled as an ODR and their health indicator is expandable to an ODR; in this case, simultaneous equations with ODR arise. Realizing the difficulty of handling simultaneity/endogeneity in multivariate models with ODR, the following papers avoided the difficulty in different ways. Carlson and Dunkelberg (1989) and McIntosh *et al.* (2000) dealt with multiple ODRs of firms (output decision, employment decision, pricing decision and so on) which are simultaneously related, but they avoided the simultaneity by assuming a recursive system. Nadeu *et al.* (1995) had simultaneous equations with ODR variables, but they treated the ODR variables as continuous. Watts and Lynch (1989) estimated independent ODR models for introductory micro- and macroeconomics course grades, but one can easily think of possible simultaneity in this case because knowledge in one subject can influence performance in the other. Machin and Stewart (1990) estimated a single-equation ODR model for a financial performance measure of firms with firm attributes as regressors, many of which are likely to be simultaneously related to the financial performance.

The rest of this paper is organized as follows. Section 2 explores the identification problem for ODR simultaneous equations with constant thresholds. Section 3 allows the thresholds to depend on regressors. Section 4 presents specification tests to be used in Section 5 which provide an empirical example using joint labour-supply decisions in farm households. Finally, Section 6 concludes. Throughout this paper, we will assume that random variables are independent and identically distributed across $i = 1, \ldots, N$, and often omit the subscript i.

2. CONSTANT-THRESHOLD ODR MODELS

Consider SF simultaneous equations with J (≥ 2) endogenous variables:

$$y_{ij}^* = \sum_{m=1, m \neq j}^{J} \alpha_{jm} y_{im}^* + x_{ij}' \beta_j + u_{ij}, \quad j = 1, \ldots, J, \tag{2}$$

where y_{ij}^* is a latent endogenous variable, x_{ij} is a $k_j \times 1$ vector of explanatory variables with 1 as its first component and $k_j \geq 2$, u_{ij} is an error term, and α_{jm} and β_j are unknown SF parameters, $j, m = 1, \ldots, J$. Let x_i denote the $k \times 1$ vector consisting of all the elements of x_{ij}, $j = 1, \ldots, J$, and assume

$$E(x_i u_{ij}) = 0 \text{ for all } j, \quad E(x_i x_i') \text{ is of full rank.}$$

Let the order of the components in x_{ij} follow that in x_i. We will refer to α_{jm} and β_j as 'endogenous SF parameter' and 'exogenous SF parameter', respectively. Section 2.1 reviews RF ODR identification, Section 2.2 shows the SF identification and Section 2.3 presents further identification results under a cross-equation restriction.

2.1. Reduced-form identification

Solving the system of equation (2) for y_{ij}^*, we get the RF equations

$$y_{ij}^* = x_i' \eta_j + v_{ij}, \quad j = 1, \ldots, J, \tag{3}$$

where η_j is an RF parameter vector and v_{ij} is an RF error. As in (1), suppose we observe, for each individual, only x_i and

$$y_{ij} = \sum_{r=1}^{R-1} 1\left[y_{ij}^* > \gamma_{jr}\right], \quad j = 1, \ldots, J; \tag{4}$$

the thresholds γ_{jr} are allowed to vary across equations in (4).

Subtract the intercept γ_{j1} from both sides of $x_i' \eta_j + v_{ij} > \gamma_{jr}$ and divide through by an unknown positive scale factor σ_j to get

$$\frac{x_i' \eta_j - \gamma_{j1}}{\sigma_j} + \frac{v_{ij}}{\sigma_j} > \frac{\gamma_{jr} - \gamma_{j1}}{\sigma_j}, \quad r = 1, \ldots, R-1.$$

The scale factor σ_j is the standard deviation (SD) of v_j when (4) is estimated by ordered probit, and it is $|\eta_{jk}|$ where η_{jk} is the kth component of η_j and η_j is estimated by a semiparametric estimator as in Melenberg and Van Soest (1996). The point is that although we will proceed mostly with parametric estimators (ordered probit), our identification results (to be presented) are also applicable for semiparametric methods.

As is well known, the identified parameters for the jth RF ODR equation are

$$\begin{aligned}
\delta_{j1} &\equiv \frac{\eta_{j1} - \gamma_{j1}}{\sigma_j}, && \text{(for the intercept)},\\
\delta_{js} &\equiv \frac{\eta_{js}}{\sigma_j}, \quad s = 2, \ldots, k, && \text{(for the slopes)},\\
\tau_{jm} &\equiv \frac{\gamma_{jm} - \gamma_{j1}}{\sigma_j}, \quad m = 2, \ldots, R-1, && \text{(for the thresholds)}.
\end{aligned} \tag{5}$$

The intercept δ_{j1} requires special care, as will be seen further on. Define

$$\delta_j \equiv (\delta_{j1}, \delta_{j2}, \ldots, \delta_{jk})', \quad \forall j \tag{6}$$

which is a $k \times 1$ vector for the identified RF ODR regression parameters.

2.2. *Structural-form identification*

Insert (3) into both sides of (2) to get

$$\begin{aligned}
x_i'\eta_j + v_{ij} &= \sum_{m=1, m\neq j}^{J} \alpha_{jm}(x_i'\eta_m + v_{im}) + x_{ij}'\beta_j + u_{ij}\\
&= x_i'\left(\sum_{m=1, m\neq j}^{J} \alpha_{jm}\eta_m\right) + x_{ij}'\beta_j + \left(\sum_{m=1, m\neq j}^{J} \alpha_{jm}v_{im} + u_{ij}\right).
\end{aligned} \tag{7}$$

Define the $k \times k_j$ 'selection matrix' S_j such that

$$\underset{1\times k_j}{x_{ij}'} = \underset{1\times k}{x_i'} \cdot \underset{k\times k_j}{S_j};$$

S_j is a known matrix consisting of 1's and 0's. With both x_i and x_{ij} having 1 as their first element, the first row of S_j is $(1, 0, \ldots, 0)\ \forall j$. For example, if $k = 3, x_i = (1, w_i, z_i)', k_j = 2$ and $x_{ij} = (1, z_i)'$, then

$$S_j = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad x_{ij}' = x_i'S_j \;\Leftrightarrow\; [1, z_i] = [1, w_i, z_i]\begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Using the definition of S_j, rewrite (7) as

$$x_i'\eta_j + v_{ij} = x_i'\left(\sum_{m=1, m\neq j}^{J} \alpha_{jm}\eta_m + S_j\beta_j\right) + \left(\sum_{m=1, m\neq j}^{J} \alpha_{jm}v_{im} + u_{ij}\right). \tag{7$'$}$$

Pre-multiply both sides by x_i and take expectations to get rid of the error terms. Since $E(x_i x_i')$ is of full rank, the resulting equation is equivalent to

$$\eta_j = \sum_{m=1, m\neq j}^{J} \alpha_{jm}\eta_m + S_j\beta_j, \quad j = 1, \ldots, J. \tag{8}$$

This equation links the RF parameters η_j to the SF parameters α_{jm} and β_j. In ODR, however, the η_j's are not fully identified. Hence, we need an equation that links the *identified* RF parameters in the ODR model, defined in (5), to the SF parameters. The equation is (10), and we show the steps leading to (10) in the following lines.

Divide (8) by σ_j to account for the up-to-scale identification of η_j:

$$\frac{\eta_j}{\sigma_j} = \sum_{m=1,m\neq j}^{J} \frac{\alpha_{jm}\eta_m}{\sigma_j} + \frac{S_j\beta_j}{\sigma_j}. \tag{8$'$}$$

To avoid the wrong normalization η_m/σ_j, rewrite this equation as

$$\frac{\eta_j}{\sigma_j} = \sum_{m=1,m\neq j}^{J} \alpha_{jm}\frac{\sigma_m}{\sigma_j}\frac{\eta_m}{\sigma_m} + \frac{S_j\beta_j}{\sigma_j}. \tag{8$''$}$$

Although the slopes in η_j are identified up to scale, the intercept in η_j is identified up to scale and location [recall (5)]. For this, define 0_m as the $m \times 1$ null vector and

$$\mu_{jm} \equiv \alpha_{jm}\frac{\sigma_m}{\sigma_j}, \quad \vartheta_1 \equiv \begin{bmatrix} 1 \\ 0_{k-1} \end{bmatrix} \quad \text{and} \quad \vartheta_{1j} \equiv \begin{bmatrix} 1 \\ 0_{k_j-1} \end{bmatrix}.$$

Note that $\vartheta_1 = S_j\vartheta_{1j}$. Subtract $\vartheta_1\gamma_{j1}/\sigma_j$ from both sides of (8$''$) to get

$$\frac{\eta_j}{\sigma_j} - \frac{\vartheta_1\gamma_{j1}}{\sigma_j} = \sum_{m=1,m\neq j}^{J} \mu_{jm}\frac{\eta_m}{\sigma_m} + \frac{S_j\beta_j}{\sigma_j} - \frac{\vartheta_1\gamma_{j1}}{\sigma_j}.$$

Doing the analogous subtraction from the intercept in each η_m/σ_m of the right-hand-side sum, we get, for all j,

$$\frac{\eta_j}{\sigma_j} - \frac{\vartheta_1\gamma_{j1}}{\sigma_j} = \sum_{m=1,m\neq j}^{J} \mu_{jm}\left(\frac{\eta_m}{\sigma_m} - \frac{\vartheta_1\gamma_{m1}}{\sigma_m}\right) + \sum_{m=1,m\neq j}^{J} \mu_{jm}\frac{\vartheta_1\gamma_{m1}}{\sigma_m} + S_j\frac{\beta_j}{\sigma_j} - \frac{\vartheta_1\gamma_{j1}}{\sigma_j} \iff$$
$$\frac{\eta_j}{\sigma_j} - \frac{\vartheta_1\gamma_{j1}}{\sigma_j} = \sum_{m=1,m\neq j}^{J} \mu_{jm}\left(\frac{\eta_m}{\sigma_m} - \frac{\vartheta_1\gamma_{m1}}{\sigma_m}\right) + S_j\frac{\beta_j}{\sigma_j} + \vartheta_1\left(\sum_{m=1,m\neq j}^{J} \mu_{jm}\frac{\gamma_{m1}}{\sigma_m} - \frac{\gamma_{j1}}{\sigma_j}\right). \tag{9}$$

Since $\vartheta_1 = S_j\vartheta_{1j}$ and $\delta_j = \eta_j/\sigma_j - \vartheta_1\gamma_{j1}/\sigma_j$, (9) becomes

$$\underset{k\times 1}{\delta_j} = \sum_{m=1,m\neq j}^{J} \underset{1\times 1}{\mu_{jm}}\,\underset{k\times 1}{\delta_m} + \underset{k\times k_j}{S_j}\left[\underset{k_j\times 1}{\frac{\beta_j}{\sigma_j}} + \underset{k_j\times 1}{\vartheta_{1j}}\left(\sum_{m=1,m\neq j}^{J} \mu_{jm}\frac{\gamma_{m1}}{\sigma_m} - \frac{\gamma_{j1}}{\sigma_j}\right)\right] \tag{10}$$

where the dimensions are provided for some terms to prevent confusion. The last term in (9) next to ϑ_1 is a scalar affecting only the intercept of β_j, and this is now made explicit with ϑ_{1j} in (10).

Since δ_j's are the identified ODR RF regression parameters and S_j 's are known, (10) shows the *identified parameters for SF j* in (2):

$$\underbrace{\mu_{jm}\ m = 1, \ldots, J,\ m \neq j}_{\text{endogenous SF parameters}}, \quad \underbrace{\frac{\beta_{j1}}{\sigma_j} + \sum_{m=1,m\neq j}^{J} \mu_{jm}\frac{\gamma_{m1}}{\sigma_m} - \frac{\gamma_{j1}}{\sigma_j}}_{\text{exogenous SF parameter (intercept)}}, \quad \underbrace{\frac{\beta_{j2}}{\sigma_j}, \ldots, \frac{\beta_{jk_j}}{\sigma_j}}_{\text{exogenous. SF parameters (slope)}}; \tag{11}$$

this shows exactly how the identification of the SF parameters in (2) is restricted. The intercept in β_j is identified up to scale and location—i.e., not identified. This is because 1 appears as a 'regressor' in both the thresholds and the regression function. When we allow for varying thresholds later on, the same reasoning will show that the regression coefficient of a regressor that appears in both the thresholds and the regression function is not identified.

To see the order and rank conditions for the identification of the SF parameter in (11), imagine a least-square estimation (LSE) of δ_j on

$$D_j \equiv (\delta_m, \ m = 1, \ldots, J, \ m \neq j, \ \ S_j) \tag{12}$$

in (10). First, note that the dimensions of (11) and D_j are, respectively, $(J - 1 + k_j) \times 1$ and $k \times (J - 1 + k_j)$. Second, the LSE requires $D_j' D_j$ to be of full rank (*rank condition*, the rank being $J - 1 + k_j$). Third, for this to hold, it is necessary to have $J - 1 + k_j \leq k \Leftrightarrow k_j \leq k - (J - 1)$: at least $J - 1$ variables in x should be excluded from SF j (*order condition*).

A literature search did not turn up any paper that presents the SF parameter identification step by step as in (7) to (10), even though it is conceptually straightforward and has been more or less shown in Lee (1995) and Kimhi and Lee (1996). Had we had observed continuous responses, we would have stopped at applying LSE to (8) to derive rank/order conditions there; the steps (7) and (8) are far simpler than the usual textbook discussion of simultaneous-equation identification with continuous responses.

2.3. Identification under a cross-equation restriction

As already mentioned, further identification of SF parameters is possible under a certain condition, as the following theorem shows:

Theorem 1. *In the constant-threshold ODR model, (2) to (4), if at least one threshold difference is the same across all RF equations and if the rank conditions hold, then the endogenous SF parameters are fully identified. Also, for a non-constant component of x_i appearing in the jth and mth SF equation as the λth exogenous regressor, the ratio $\beta_{j\lambda}/\beta_{m\lambda}$ is identified where λ can be any integer in* $[2, min(k_j, k_m)]$.

Proof. Suppose $R \geq 4$, and without loss of generality, suppose that $\gamma_{j3} - \gamma_{j2}$ is the same unknown constant, say q_{32}, $\forall j$. Then, by the definitions in (5),

$$\tau_{j3} - \tau_{j2} = \frac{\gamma_{j3} - \gamma_{j2}}{\sigma_j} = \frac{q_{32}}{\sigma_j}. \tag{13}$$

Thus, for $j, m = 1, \ldots, J$, recalling $\mu_{jm} \equiv \alpha_{jm} \cdot \sigma_m/\sigma_j$,

$$\frac{\sigma_j}{\sigma_m} = \frac{\tau_{m3} - \tau_{m2}}{\tau_{j3} - \tau_{j2}} \implies \alpha_{jm} = \mu_{jm} \cdot \frac{\tau_{m3} - \tau_{m2}}{\tau_{j3} - \tau_{j2}}. \tag{14}$$

Since μ_{jm} and $(\tau_{m3} - \tau_{m2})/(\tau_{j3} - \tau_{j2})$ are identified, α_{jm} is identified. As for $\beta_{j\lambda}/\beta_{m\lambda}$, it can be written as an identified form

$$\frac{\beta_{j\lambda}}{\beta_{m\lambda}} = \frac{\beta_{j\lambda}/\sigma_j}{\beta_{m\lambda}/\sigma_m} \cdot \frac{\tau_{m3} - \tau_{m2}}{\tau_{j3} - \tau_{j2}}. \tag{15}$$

If $R = 3$, assume that $\gamma_{j2} - \gamma_{j1}$ is the same unknown constant $\forall j$, and the preceding proof goes through with $\tau_{j3} - \tau_{j2}$ and $\tau_{m3} - \tau_{m2}$ replaced by τ_{j2} and τ_{m2}, respectively. □

If conditions such as (13) hold for other threshold differences, say $\gamma_{j2} - \gamma_{j1}$, then this may enhance the efficiency of the estimator in use for the SF parameters but not the identification. Nevertheless, as will be shown in Section 4, if both $\gamma_{j3} - \gamma_{j2} = q_{32}$ and $\gamma_{j2} - \gamma_{j1} = q_{21}$ hold $\forall j$, then the cross-equation restriction is testable. The part of Theorem 1 dealing with $\beta_{j\lambda}/\beta_{m\lambda}$ helps compare the coefficients of the same regressor across SF equations. Theorem 1 stands even if the number of ODR categories varies across equations.

In Watts and Lynch (1989), if course grades have the same threshold differences, then (13) holds. In Kimhi and Lee (1996), each ODR labour-supply equation has four categories ($R = 4$): no work, 1/3 part-time, 2/3 part-time and full-time work. The model consists of four equations ($J = 4$): two for farm labour supply (male and female) and two for off-farm labour supply (male and female). The terms 'part-time' and 'full-time' are not clearly defined in the data source. If the two spouses share the same definitions of part-time and full-time, then (13) holds.

Essentially, (14) states that the unidentified scale difference between σ_j and σ_m is captured by the identified threshold differences $\tau_{j3} - \tau_{j2}$ and $\tau_{m3} - \tau_{m2}$. If $\tau_{j3} - \tau_{j2}$ happens to be the interquartile range for the y_j^* distribution, $\tau_{j3} - \tau_{j2}$ is as legitimate as σ_j in representing the scale of y_j^*; in fact, any difference between two location parameters can be used as a scale measure. Thus, the assumption of the same threshold differences across equations is comparable to that of the same scales across the equations, and with this, the endogenous SF parameters are fully identified as if y_j^* had been fully observed.

3. VARYING-THRESHOLD ODR MODELS

In applying ODR models with unknown thresholds, one often confronts the question of threshold constancy: Given that the regression function depends on x_i, would it not be likely that the thresholds are functions of x_i as well: Terza (1985), for example, used a bond-rating ODR variable: the bond-rating companies may not be applying the same standards (the thresholds) to all companies, and indeed this was found to be the case. As another example, suppose that y^* is worker promotability and y is the observed rank. If the thresholds depend on race (or sex), then this is evidence of discrimination in promotion. Winter-Ebmer and Zweimuller (1997) and Pudney and Shields (2000) applied an ODR model with varying thresholds to promotion processes. However, the identification issue has never been dealt with adequately in the literature. In this section, we allow the thresholds to depend on regressors and see how this affects the SF identification for simultaneous equations in ODR. Section 3.1 reviews RF ODR identification, Section 3.2 shows the SF identification and Section 3.3 presents further identification results under a cross-equation restriction. Towards the end, Section 3.3 also reviews the structural-equation and errors-in-variable approach, which is popular in social science disciplines other than economics to show how this approach and ours are related.

3.1. Reduced-form identification

The most general setup for varying thresholds would allow for different covariates across the regression functions and thresholds. However, to simplify our discussion, we assume that all thresholds in all equations share the same $c \times 1$ vector of covariates z_i; as will be seen shortly, our findings are already complicated, even under this simplification; a further generalization is left for future research. Since the first threshold is subtracted from the regression function, any regressor in the thresholds should also appear in the regression function. Without loss of generality, let z_i be the first $c \times 1$ sub-vector of x_i ($c \leq k$).

Define the variables in x_i other than z_i as $\tilde{x}_i$ and define η_{zj} and $\tilde{\eta}_j$ as the parameter vectors for z_i and $\tilde{x}_i$, respectively, in the jth RF equation:

$$x_i = \begin{bmatrix} z_i \\ \tilde{x}_i \end{bmatrix} \quad \text{and} \quad \eta_j = \begin{bmatrix} \eta_{zj} \\ \tilde{\eta}_j \end{bmatrix} \quad \forall j;$$

if $z_i = x_i$, then $\tilde{x}_i$ and $\tilde{\eta}_j$ should be removed. Let θ_{jr} be the coefficient vector for threshold r in RF equation j: that is,

$$\gamma_{ijr} \equiv z_i'\theta_{jr}, \quad r = 1, \ldots, R-1; \tag{16}$$

as mentioned above, the most general model would have $z_{ijr}'\theta_{jr}$ on the right-hand side of (16). Although we will consider thresholds depending only on the observed x_i, the thresholds in RF j can, in fact, share a common error term ω_{ij} as in $\gamma_{ijr} = z_i'\theta_{jr} + \omega_{ij}$. In this case, as γ_{ij1} is subtracted from the regression function and the other thresholds, ω_{ij} is absorbed into v_{ij} and the threshold differences become free of ω_{ij}. What we do not allow in this paper is for ω_{ij} to carry different coefficients for different thresholds: that is, $\gamma_{ijr} = z_i'\theta_{jr} + \psi_{jr}\,\omega_{ij}$ is not allowed where ψ_{jr} is the coefficient for equation j and threshold r.

The ODR rule for RF equation j with varying thresholds is

$$y_{ij} = \sum_{r=1}^{R-1} 1[x_i'\eta_j + v_{ij} > \gamma_{ijr}]. \tag{17}$$

To guarantee that $\gamma_{ij1} <, \cdots, < \gamma_{ij,R-1}\,\forall i, j$, assume that $\theta_{j1} <, \cdots, < \theta_{j,R-1}\,\forall j$, and linearly transform z_i such that $z_i > 0$ for all i. There is no loss of generality here as long as each component of z_i affects all thresholds in the same direction; for instance, if a positive component z_{ic} of z_i affects all thresholds negatively, then we can use $-z_{ic} + M_c$ instead of z_{ic} where M_c is chosen such that $-z_{ic} + M_c > 0\,\forall i$.

Subtract $z_i'\,\theta_{j1}$ and divide through by σ_j in (17) to get

$$y_{ij} = \sum_{r=1}^{R-1} 1\left[\frac{z_i'(\eta_{zj} - \theta_{j1})}{\sigma_j} + \frac{\tilde{x}_i'\tilde{\eta}_j}{\sigma_j} + \frac{v_{ij}}{\sigma_j} > \frac{z_i'(\theta_{jr} - \theta_{j1})}{\sigma_j}\right]. \tag{18}$$

The identified parameters in RF j are [compare to (5)]

$$\begin{aligned} \delta_{zj} &\equiv \frac{\eta_{zj} - \theta_{j1}}{\sigma_j} \qquad (\text{for } z_i), \\ \tilde{\delta}_j &\equiv \frac{\tilde{\eta}_j}{\sigma_j} \qquad (\text{for } \tilde{x}_i), \\ \frac{\theta_{jr} - \theta_{j1}}{\sigma_j}&, \quad r = 2, \ldots, R-1, \quad (\text{for the thresholds}); \end{aligned} \tag{19}$$

let $\delta_j \equiv (\delta_{zj}', \tilde{\delta}_j')'$.

3.2. Structural-form identification

Denote column m of the identity matrix I_k as e_m. Then the selection matrix S_j consists of e_ms. Define the 'threshold-variable' selection matrix L_c with

$$z_i' = x_i' \cdot [e_1, \dots, e_c] = x_i' L_c. \tag{20}$$

Recall that, for constant thresholds, the intercept in β_j is not identified because 1 appears in the thresholds. For varying thresholds, the components of β_j for the regressors appearing in both x_{ij} and z_i are not identified. Instead of (9), we get [recall (19)]

$$\delta_j = \sum_{m=1, m\neq j}^{J} \mu_{jm}\delta_m + S_j \frac{\beta_j}{\sigma_j} + L_c\left(\sum_{m=1, m\neq j}^{J} \mu_{jm}\frac{\theta_{m1}}{\sigma_m} - \frac{\theta_{j1}}{\sigma_j}\right). \tag{21}$$

The expression corresponding to (10) is somewhat complicated due to (non-) overlapping columns in S_j and L_c. We show two examples first, and then a general formula in (24).

Suppose $J = 2, k = 4, k_j = 3, c = 2$, and the first, second, and fourth elements of x_i appear in SF j, carrying the coefficients β_{j1}, β_{j2} and β_{j3}, respectively. Since μ_{jm} is a scalar and θ_{m1} and θ_{j1} are 2×1 vectors, define the vector to the right of L_c in (21) as $(\xi_1, \xi_2)'$. Then, S_j and the terms to the right of S_j in (21) are

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\beta_{j1}}{\sigma_j} \\ \frac{\beta_{j2}}{\sigma_j} \\ \frac{\beta_{j3}}{\sigma_j} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\beta_{j1}}{\sigma_j} + \xi_1 \\ \frac{\beta_{j2}}{\sigma_j} + \xi_2 \\ \frac{\beta_{j3}}{\sigma_j} \end{bmatrix}. \tag{22}$$

Here, β_{j3} is identified up to σ_j while β_{j1} and β_{j2} are not.

Suppose everything is the same but now $c = 4$. Since θ_{m1} and θ_{j1} are 4×1 vectors, define the vector to the right of L_c in (21) as $(\xi_1, \xi_2, \xi_3, \xi_4)'$. Then S_j and the terms to the right of S_j in (21) are

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\beta_{j1}}{\sigma_j} \\ \frac{\beta_{j2}}{\sigma_j} \\ \frac{\beta_{j3}}{\sigma_j} \end{bmatrix} + I_4 \cdot \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{bmatrix} = I_4 \cdot \begin{bmatrix} \frac{\beta_{j1}}{\sigma_j} + \xi_1 \\ \frac{\beta_{j2}}{\sigma_j} + \xi_2 \\ \xi_3 \\ \frac{\beta_{j3}}{\sigma_j} + \xi_4 \end{bmatrix}. \tag{23}$$

Here, no component of β_j is identified up to scale.

What (22) and (23) show is that the last two terms in (21) involving S_j and L_c can be written as a single term—as in the right-hand sides of (22) and (23)—which is a product of two matrices, say $Q_j \cdot \zeta_j$, where the columns of Q_j are the 'union' of the columns of S_j and L_c and ζ_j is defined accordingly, as (22) and (23) illustrate. Let Q_j be the column dimension of Q_j; then Q_j is $k \times q_j$

and ζ_j is $q_j \times 1$. The expression corresponding to (10) is

$$\underset{k\times 1}{\delta_j} = \sum_{m=1,m\neq j}^{J} \underset{1\times 1}{\mu_{jm}} \underset{k\times 1}{\delta_m} + \underset{k\times q_j}{Q_j} \underset{q_j\times 1}{\zeta_j} . \tag{24}$$

The $(J-1)+q_j$ identified SF parameters in (24) are

$$(\mu_{jm},\ m=1,\ldots,J,\ m\neq j,\ \zeta_j')'. \tag{25}$$

Comparing (10) and (24), two findings emerge. First, endogenous SF parameter identification is the same for both constant and varying thresholds, subject to a change in rank condition as shown in the following paragraph; this is because the endogenous regressors are excluded from the thresholds. Second, exogenous SF parameter identification is different: S_j appears in (10) while Q_j appears in (24). That is, the identification of β_j depends on the overlap between x_{ij} and z_i. One extreme case is $c = 1$ (constant threshold), where all elements of β_j except the intercept are identified up to σ_j. The other extreme case is $z_i = x_{ij}$ (i.e. $c = k$), where no component of β_j is identified up to σ_j as (23) shows.

For the order and rank conditions for the SF parameters in (25), imagine an LSE for (24) of δ_j on $F_j \equiv (\delta_m, m=1,\ldots,J, m\neq j, Q_j)$. First, note that the dimensions of (25) and F_j are $(J-1+q_j)\times 1$ and $k\times(J-1+q_j)$, respectively. Second, the LSE requires $F_j'F_j$ to be of full rank (*rank condition*, the rank being $J-1+q_j$). Third, for this to hold, it is necessary to have (*order condition*):

$$J-1+q_j \leq k \ \Leftrightarrow\ q_j \leq k-(J-1). \tag{26}$$

To better understand (26), consider the following cases. First, $q_j = k_j$ so that z_i is properly included in x_{ij} as in (22); here, the order condition is $k_j \leq k-(J-1)$: at least $J-1$ variables in x_i should be excluded from SF j, which is analogous to the constant-threshold case. The first c-many elements of β_j are not identified, while the rest of β_j are identified up to σ_j. Second, $q_j = c$, so x_{ij} is included in z_i as in (23); here, the order condition is $c \leq k-(J-1)$: at least $J-1$ variables in x_i should be excluded from z_i. No component of β_j is identified up to σ_j, as (23) illustrates.

3.3. Identification under a cross-equation restriction

We now present Theorem 2, which is analogous to Theorem 1. As for the identification of β_j's, we only deal with the case of z_i properly included in x_{ij}, which parallels the identification of β_j's with constant thresholds.

Theorem 2. *In the varying-threshold ODR model (16) and (17), if at least one threshold difference is the same across all RF equations, and if the rank conditions hold, then the endogenous SF parameters are fully identified; the same threshold difference may depend on i. Also, for a non-constant component of x_i appearing in the jth and mth SF equation as the λth exogenous regressor, if min$(k_j, k_m) > c$, the ratio $\beta_{j\lambda}/\beta_{m\lambda}$ is identified, where λcan be any integer in* $[c+1,$ *min*$(k_j, k_m)]$.

Proof. Let

$$\tau_{ijr} = \frac{\gamma_{ijr}}{\sigma_j} = \frac{z_i'\theta_{jr}}{\sigma_j}. \tag{27}$$

Suppose, for some unknown q_{32i},

$$\tau_{ij3} - \tau_{ij2} = \frac{\gamma_{ij3} - \gamma_{ij2}}{\sigma_j} = \frac{z_i'(\theta_{j3} - \theta_{j2})}{\sigma_j} = \frac{q_{32i}}{\sigma_j} \ \forall j. \tag{28}$$

The rest of the proof is analogous to that for Theorem 1. □

In social science disciplines such as psychology, sociology and education, 'SF equations with unobserved factors' are often used. Having presented all of our identification results, here we compare our models to SF equations with unobserved factors. To simplify the exposition, ignore the regressors for a while and assume that the expected values of all random variables are 0. Consider the stacked latent SF equations

$$y^* = Ay^* + u \implies y^* = (I_J - A)^{-1}u, \quad \text{where } y^* \equiv (y_1^*, \ldots, y_J^*)', \ u \equiv (u_1, \ldots, u_J)', \tag{29}$$

and A is a conformably defined endogenous SF parameter matrix such that $(I_J - A)^{-1}$ exists. Suppose that what is observed is not y^* but a $K \times 1$ vector z^* such that

$$\underset{K\times 1}{z^*} = \underset{K\times J}{C} \ \underset{J\times 1}{y^*} + \underset{K\times 1}{\varepsilon}, \tag{30}$$

where C is a parameter matrix and ε is an error vector. If $K > J$, then the J-many latent variables (or 'factors') are behind the K-many manifested variables; (30) is a 'factor (analysis) model'. This is plausible, for example, if y^* is various underlying ability indices and z^* is continuously distributed exam scores. Equation (30) is also an 'errors-in-variable' model; this name is justified even if $K = J$.

Equations (29) and (30) are called a 'structural equation model (SEM)' in the social science disciplines, and the SEM includes the usual simultaneous equations in econometrics (with x added) as special cases. In most cases, the SEM assumes normality for the error terms and uses the variances of the observed variables to identify the parameters. Despite the generality of the SEM, however, this entails a severe identification problem. To see this, assume $u \sim N(0, \Omega_u)$, $\varepsilon \sim N(0, \Omega_\varepsilon)$, and independence between u and ε to get

$$E(y^*y^{*\prime}) = (I_J - A)^{-1}\Omega_u(I_J - A')^{-1} \implies E(z^*z^{*\prime}) = C(I_J - A)^{-1}\Omega_u(I_J - A')^{-1}C' + \Omega_\varepsilon.$$

There are $K(K+1)/2$ identified entities on the left-hand side, whereas there are as many as

$$K \times J \text{ (from } C) + J \times J \text{ (from } A) + \frac{J(J+1)}{2} \text{ (from } \Omega_u) + \frac{K(K+1)}{2} \text{ (from } \Omega_\varepsilon) \tag{31}$$

parameters at maximum on the right-hand side. Adding the $k \times 1$ regressor vector x to (29), (31) increases by $J \times k$. For example, if $K = J = 4$, then only 10 parameters are identified in $E(z^*z^{*\prime})$ whereas (31) is $16 + 16 + 10 + 10 = 46$.

Turning to the ODR version of the SEM, now suppose $z_i = \sum_{r=1}^{R-1} 1[z_i^* > \gamma_r]$ is observed instead of z_i^*. Using the joint distribution of z obtained under the normality assumptions, the correlation matrix for z^* is identified along with the thresholds γ_r's. This means K-fewer identified entities for (31), because the scale information of z^* is lost in z. Allowing for x to enter the thresholds may be done at this ODR stage although we could not find any study in the SEM literature that does this.

In principle, the SEM includes our ODR simultaneous equations as a special case when $z^* = y^*$. But the estimation approaches as reviewed in chapters 8 and 11 of Wansbeek and Meijer (2000) are not helpful for our model because they are geared for errors-in-variable models; moreover, they do not consider regressor-dependent thresholds or cross-equation restrictions for threshold differences. Finally, estimating the correlation matrix across J-many RF equations would not be an easy task, to say the least.

To avoid the kind of identification problems as in (31), econometricians invoke a priori restrictions. But as can be seen in our empirical example, such restrictions are imposed with due justification. The identification problem and ad hoc solutions invoked in practice for the SEM are so frustrating that Bartholomew and Knott (1999) state, in the last sentence of their book, that "*When we come to models for relationships between latent variables we have reached a point where so much has to be assumed that one might justly conclude that the limits of scientific usefulness have been reached if not exceeded*." It seems that either the SF approach or the errors-in-variable approach is doable, but not both jointly.

4. SPECIFICATION TESTS

In this section, first, the log likelihood and score functions for the constant- and varying-threshold probits are shown to implement the likelihood-ratio and score tests for threshold constancy. Second, a simple Wald test for threshold constancy is introduced that requires only the constant-threshold model estimates; a small-scale simulation study comparing the score and Wald tests is also presented. Third, tests for the cross-equation threshold-difference constancy are proposed. All tests in this section are for RF models, not SF.

Since examining one RF equation is enough, we drop the subscript j and consider (1) with v_i following $N(0, \sigma^2)$ independently of x_i. The identified parameters for the Constant-Threshold Probit (CPRO) are [recall (5) and (6) with j removed]

$$\delta \equiv (\delta_1, \delta_s')', \ \tau_m \equiv \frac{\gamma_m - \gamma_1}{\sigma}, \ m = 2, \ldots, R-1 \ \text{ where } \delta_1 \equiv \frac{\eta_1 - \gamma_1}{\sigma}, \ \delta_s \equiv \left(\frac{\eta_2}{\sigma}, \ldots, \frac{\eta_k}{\sigma}\right)';$$

note that $t_0 = \tau_0 = -\infty, t_1 = \tau_1 = 0$ and $t_R = \tau_R = \infty$. Define

$$y_{ir} = 1 \text{ if } y_i = r \text{ and } 0 \text{ otherwise}, \quad r = 0, 1, \ldots, R-1.$$

Let Φ and ϕ denote the $N(0, 1)$ distribution function and its density. The CPRO maximizes

$$Q_c(d, t) \equiv \sum_{i=1}^{N} \sum_{r=1}^{R} y_{i,r-1} \ln\{\Phi(t_r - x_i'd) - \Phi(t_{r-1} - x_i'd)\} \tag{32}$$

for d and $t \equiv (t_2, \ldots, t_{R-1})'$ where d and t are estimators for δ and τ_r's, respectively.

Now consider the Varying-Threshold Probit (VPRO) with the identified parameters [recall (19) with j dropped]

$$\delta \equiv (\delta_z', \tilde{\delta}')', \ \tau_r \equiv \frac{\theta_r - \theta_1}{\sigma}, \quad r = 2, \ldots, R-1 \ \text{ where } \delta_z \equiv \frac{\eta_z - \theta_1}{\sigma}, \ \tilde{\delta} \equiv \frac{\tilde{\eta}}{\sigma};$$

δ and τ_r's for the VPRO are different from those for the CPRO, although the same notations are used. The maximand for d and $t = (t_2', \ldots, t_{R-1}')'$ is

$$Q_v(d, t) \equiv \sum_{i=1}^{N} \sum_{r=1}^{R} y_{i,r-1} \ln\{\Phi(z_i't_r - x_i'd) - \Phi(z_i't_{r-1} - x_i'd)\} \tag{33}$$

The first derivatives are

$$\frac{\partial Q_v(d,t)}{\partial d} = \sum_{i=1}^{N}\sum_{r=1}^{R} y_{i,r-1}\left\{\frac{\phi(z_i' t_r - x_i' d) - \phi(z_i' t_{r-1} - x_i' d)}{\Phi(z_i' t_r - x_i' d) - \Phi(z_i' t_{r-1} - x_i' d)}\right\} \cdot (-x_i),$$

$$\frac{\partial Q_v(d,t)}{\partial t_r} = \sum_{i=1}^{N}\sum_{r=1}^{R} \phi(z_i' t_r - x_i' d)\left\{\frac{y_{i,r-1}}{\Phi(z_i' t_r - x_i' d) - \Phi(z_i' t_{r-1} - x_i' d)}\right.$$
$$\left. - \frac{y_{ir}}{\Phi(z_i' t_{r+1} - x_i' d) - \Phi(z_i' t_r - x_i' d)}\right\} \cdot z_i. \qquad (34)$$

The derivatives for the CPRO can be obtained by simply removing the last x_i (z_i) from $\partial Q_v(d, t)/\partial d(\partial Q_v(d, t)/\partial t_r)$ and replacing $z_i' t_r$ with t_r. Instead of (34), numerical derivatives may be used in numerical optimizations.

With '$\rightsquigarrow$' denoting convergence in law, the score and likelihood ratio tests are

$$\frac{1}{\sqrt{N}}\sum_i s_{vi} \cdot \left(\frac{1}{N}\sum_i s_{vi} s_{vi}'\right)^{-1} \frac{1}{\sqrt{N}}\sum_i s_{vi}' \rightsquigarrow \chi^2_{(c-1)(R-2)}, \qquad (35)$$

$$2\{Q_v(\text{VPRO}) - Q_c(\text{CPRO})\} \rightsquigarrow \chi^2_{(c-1)(R-2),}, \qquad (36)$$

where s_{vi} denotes the ith datum score function for the VPRO [i.e., the summand in (34)] evaluated at the CPRO estimates. The score test is simpler to use, since it does not require the VPRO estimates. Surprisingly, there is a test for threshold constancy that does not even require the VPRO score functions; the test is introduced in the following.

Suppose the ODR categories are collapsed such that $1, \ldots, R-1$ are recorded as one to yield a Binary Probit (BPRO). In the BPRO, since the thresholds are non-existent, there is no issue of threshold specification. If all assumptions for the CPRO hold, then the CPRO is the efficient estimator whereas the BPRO is not; both are consistent. If the thresholds are regressor dependent, then both are inconsistent because the first threshold is subtracted from the regression function. But, since the CPRO misspecifies all of the other thresholds whereas the BPRO does not, the two estimators are inconsistent to different extents in general. This means that a Wald test for threshold constancy can be devised comparing the CPRO and BPRO.

To obtain the asymptotic variance matrix of the Wald test, define the ith datum score function for the slope coefficients δ_s in the CPRO as s_{si}, and define s_{ni} as the score function for the other parameters in the CPRO. Denoting the estimator for δ_s as $\check{\delta}_s$, it holds that

$$\sqrt{N}(\check{\delta}_s - \delta_s) = \frac{1}{\sqrt{N}}\sum_i \left\{E\left(s_s^* s_s^{*\prime}\right)\right\}^{-1} s_{si}^* + o_p(1) \equiv \frac{1}{\sqrt{N}}\sum_i \lambda_i + o_p(1), \qquad (37)$$

where s_{si}^* is the 'effective score' for the ith datum obtained by regressing s_{si} on s_{ni}:

$$s_{si}^* = s_{si} - \left(\sum_i s_{si} s_{ni}'\right)\left(\sum_i s_{ni} s_{ni}'\right)^{-1} s_{ni}.$$

Analogously, denoting the slope estimator for the BPRO as $\hat{\delta}_s$, it holds that

$$\sqrt{N}(\hat{\delta}_s - \delta_s) = \frac{1}{\sqrt{N}}\sum_i \left\{E\left(\vartheta_s^* \vartheta_s^{*\prime}\right)\right\}^{-1} \vartheta_{si}^* + o_p(1) \equiv \frac{1}{\sqrt{N}}\sum_i \psi_i + o_p(1), \qquad (38)$$

where ϑ^*_{si} is the effective score for the ith datum obtained analogously to s^*_{si}. Thus we have, under the null hypothesis of constant thresholds,

$$\sqrt{N}(\check{\delta}_s - \hat{\delta}_s) = \frac{1}{\sqrt{N}} \sum_i (\lambda_i - \psi_i) + o_p(1) \rightsquigarrow N(0, \Omega_{\lambda\psi}), \quad \text{where}$$
$$\Omega_{\lambda\psi} = N^{-1} \sum_i (\hat{\lambda}_i - \hat{\psi}_i)(\hat{\lambda}_i - \hat{\psi}_i)' + o_p(1);$$

$\hat{\lambda}_i - \hat{\psi}_i$ is obtained by replacing the parameters in $\lambda_i - \psi_i$ with consistent estimates. Thus the Wald test statistic is

$$N(\check{\delta}_s - \hat{\delta}_s)' \left\{ N^{-1} \sum_i (\hat{\lambda}_i - \hat{\psi}_i)(\hat{\lambda}_i - \hat{\psi}_i)' \right\}^{-1} (\check{\delta}_s - \hat{\delta}_s) \rightsquigarrow \chi^2_{k-1}. \tag{39}$$

We could have included the intercept difference in (39) but did not because ODR can be collapsed into a binary response in different ways that result in different intercepts.

Just to verify that the score (SCORE) and Wald tests (WALD) indeed work in small samples, we conduct a brief simulation study. The null (i.e. constant threshold) model is

$$y^*_i = 1 + 1 \cdot x_i + v_i, \quad x \sim U[0, 2], \ v \sim N[0, 1],$$
$$\gamma_1 = 1, \ \gamma_2 = 2 \ (y \text{ takes } 0, 1, 2), \quad N = 50, 100, \quad \text{Reps} = 500,$$

and the simulation results are provided in Table 1 in two designs. For each design, the average p-value of WALD and SCORE, the proportion of the times when the p-value is smaller than 10% and the average bias of the slope estimates in the CPRO and BPRO are reported to show how large a difference between the CPRO and BPRO it takes to get the p-value for WALD.

In the first design where the CPRO holds and is efficient, SCORE does better than WALD in terms of the average p-value, but SCORE's actual level falls much below the nominal level 10%. SCORE is thus likely to have the lower power. In the second design, the threshold is regressor dependent with $\gamma_2 = 1 + 1.5x$. Both tests seem to have good powers, with WALD performing somewhat better. The bias column shows the substantial biases driving WALD.

Turning to testing for the cross-equation threshold-difference constancy for Theorem 1 with CPRO, the restriction is not testable if there are only three categories. But *if there are at least four categories*, then the following strengthened version

$$\gamma_{j3} - \gamma_{j2} = q_{32}, \ \text{ and } \ \gamma_{j2} - \gamma_{j1} = q_{21} \ \ \forall j \tag{40}$$

is testable with its implications (recall $\gamma_{j1} = 0$)

$$\frac{\tau_{j3} - \tau_{j2}}{\tau_{j2}} = \frac{\tau_{j3}}{\tau_{j2}} - 1 = \frac{q_{32}}{q_{21}} \ \forall j \implies \frac{\tau_{m3}}{\tau_{m2}} = \frac{\tau_{j3}}{\tau_{j2}} \ \ \forall j \neq m. \tag{41}$$

Table 1. Score and Wald tests.

	N	Avg.PV-Wald (<10%)	Aavg.PV-Score (<10%)	BIAS (CPRO, BPRO)
Const. thresh.	50	0.492 (0.104)	0.582 (0.054)	0.044, 0.044
	100	0.473 (0.106)	0.563 (0.024)	0.014, 0.013
Varying thresh.	50	0.009 (0.990)	0.031 (0.920)	−0.895, −1.521
	100	0.000 (1.000)	0.002 (1.000)	−0.926, −1.535

Since our data set has four categories, this test is applicable.

Specifically, there are (J=) four SF equations in our data, but only two are identified, as the other two do not have sufficient exclusion restrictions. Hence, we can compare the ratio of the two identified thresholds. The resulting test statistic's asymptotic distribution may be derived using the effective score functions, but such an asymptotic test is not invariant to equivalent transformations—for example, (41) can be written as $\tau_{m3}\tau_{j2} = \tau_{m2}\tau_{j3}$ (or any power function of this)—and could be unstable in finite samples depending on which form is tested. Instead, we will use a bootstrap percentile method to re-sample from the original sample with replacement to estimate $(\tau_{m3}/\tau_{m2}) - (\tau_{j3}/\tau_{j2})$. Repeating this, say 500 times, yields 500 pseudoestimates for $(\tau_{m3}/\tau_{m2}) - (\tau_{j3}/\tau_{j2})$, and the lower 2.5% and upper 2.5% quantiles of the pseudo-estimates give a bootstrap 95% confidence interval for $(\tau_{m3}/\tau_{m2}) - (\tau_{j3}/\tau_{j2})$. If the interval does not include 0, then (40) is rejected.

Testing for the cross-equation threshold-difference constancy with VPRO does not need four categories. To see why, observe that, to identify σ_m/σ_j with $(\tau_{ij3} - \tau_{ij2})/(\tau_{im3} - \tau_{im2})$ in VPRO, we need the threshold-difference ratio to be free of i, which holds iff $\theta_{j3} - \theta_{j2} = \theta_{m3} - \theta_{m2}$. In this case, there are c-many ratios available:

$$\left(\frac{(\theta_{j3} - \theta_{j2})^{(1)}/\sigma_j}{(\theta_{m3} - \theta_{m2})^{(1)}/\sigma_m}, \ldots, \frac{(\theta_{j3} - \theta_{j2})^{(c)}/\sigma_j}{(\theta_{m3} - \theta_{m2})^{(c)}/\sigma_m}\right) = \frac{\sigma_m}{\sigma_j} \cdot \underset{c\text{-many}}{(1, \ldots, 1)} \tag{42}$$

where the superscripts (1) and (c) denote the first and last component, respectively. Since there are c-many ratios, one can do a minimum distance estimation (MDE; see Lee (2002) for example) to combine the more than enough estimates for the scalar σ_m/σ_j (we owe this testing idea to a careful referee.). The (efficient) MDE is an weighted average of the ratios, and the inefficient but simpler MDE—the 'equally-weighting' MDE—is the simple average. The minimized quadratic distance in the MDE is a χ^2_{c-1} over-identification test statistic for (42).

Finding the asymptotic distribution for the over-identification test can be done with the effective score functions, but this is rather complicated. Doing a bootstrap for the over-identification test is also troublesome, because imposing the over-identification condition (42) on the pseudo data in the bootstrap is difficult. Instead, a practical procedure would be to select two significant components of z, say, the $(c-1)$th and cth element of z to test

$$\frac{(\theta_{j3} - \theta_{j2})^{(c-1)}/\sigma_j}{(\theta_{m3} - \theta_{m2})^{(c-1)}/\sigma_j} = \frac{(\theta_{j3} - \theta_{j2})^{(c)}/\sigma_j}{(\theta_{m3} - \theta_{m2})^{(c)}/\sigma_m}$$

following the analogous bootstrap percentile method as done for (41).

5. EMPIRICAL EXAMPLE

Our data set is derived from a 1995 survey of farm households in Israel, and includes 1337 "clean" observations. More details about this data set can be found in Kimhi (2004). The model pertains to the joint time-allocation decisions of farm couples, including four endogenous variables in four categories: the percentage breakdown for each y_j is given in Table 2

Each y_j has four categories, 0, 1, 2, 3, with 0 being not working and 3 being working full time. Unfortunately, both the number of equations and the number of categories for each ODR equation are four, which could be confusing. The ODR-category frequencies vary greatly across males and females. Only 17% of the males do not work on the farm, while more than half of the

Table 2. Percentage breakdown for each y_j.

		0	1	2	3	Total
y_1	Male farm labour	17	18	14	51	100
y_2	Male market labour	52	8	8	32	100
y_3	Female farm labour	52	21	14	13	100
y_4	Female market labour	42	8	15	35	100

females do not. More females work off-farm than males, but the difference is smaller than the difference in farm work.

The list of regressors is as follows (with sample means in parentheses):

age: age of husband divided by 10 (5.08)
age2: age (of husband divided by 10) squared (27.03)
cap: ln(farm capital stock+1) [in NIS 1000] (4.43)
catt: dummy for raising cattle or other livestock excluding poultry (0.08)
ed1h: dummy for male's high-school completion (0.61)
ed1c: dummy for male's education being more than high school (0.13)
ed2h: dummy for female's high school completion (0.60)
ed2c: dummy for female's education being more than high school (0.14)
land: ln(landholdings+1) [in dunams (0.23 acre)] (3.26)
nkid: number of children (age under 15) in the household (1.55)
nado: number of adolescents (age 15-21) in the household (0.88)
nadu: number of adults (age above 21) in the household (3.17)
is1: dummy for males born in Israel (0.52)
aa1: dummy for males of Asian or African origin (0.29)
is2: dummy for females born in Israel (0.56)
aa2: dummy for females of Asian or African origin (0.35)

The female's age is not included since it is highly correlated (94%) with male's age. The dummy for raising livestock is to control for the different labour demands on livestock farms. Children are considered dependents in the family, whereas adolescents and adults can be either dependents or helping hands (nado and nadu include the couple). Schooling is perceived to enhance labour productivity in both farm work and off-farm work, and dummies for being born in Israel and being of Asian/African origin are supposed to capture differences in general human capital and perhaps also labour-market discrimination.

The CPRO for the four RF equations gives the following threshold estimates:

	y_1	y_2	y_3	y_4
τ_{j2}	0.957	0.290	0.629	0.246
τ_{j3}	1.540	0.591	1.213	0.679
$\tau_{j3} - \tau_{j2} = (\gamma_{j3} - \gamma_{j2})/\sigma_j$	0.583	0.301	0.584	0.433

Across the four equations, τ_{j2} and τ_{j3} vary considerably, but there is much lower variability in $\tau_{j3} - \tau_{j2}$; particularly, $\tau_{j3} - \tau_{j2}$ for the two farm labour supplies y_1 and y_3 are almost identical.

To estimate the SF equations, we apply an MDE as in Lee (1995). The order condition requires at least three excluded exogenous variables from each SF equation. There are two justifications for these variable exclusions. First, farm variables (cap, catt and land) are unlikely to affect the market labour supplies y_2 and y_4 directly. Second, education and ethnicity of one person are unlikely to be directly relevant to the labour-supply decisions of the spouse. These considerations suggest the following list of excluded variables:

y_1	ed2h, ed2c, is2, aa2
y_2	ed2h, ed2c, is2, aa2, cap, catt, land
y_3	ed1h, ed1c, is1, aa1
y_4	ed1h, ed1c, is1, aa1, cap, catt, land

This shows that the order conditions for y_1 and y_3 are barely enough, whereas those for y_2 and y_4 are relatively plentiful. We estimated all four SF equations, but the instruments for the y_1 and y_3 SF equations were too weak collectively to give any significant results. In the following, we examine the results only for the y_2 and y_4 (off-farm work) SF equations.

The left half of Table 3 presents the CPRO y_2 and y_4 SF estimates. The μ-columns show the SF estimates, where the magnitude of each estimate is not interpretable. Applying the threshold constancy test (41) with $j = 2$ and $m = 4$, we computed $(\tau_{43}/\tau_{42}) - (\tau_{23}/\tau_{22})$ using 500 pseudo-samples and obtained the bootstrap 95% confidence interval (0.311, 1.168). Since this interval does not include zero, (41) is rejected. We will, however, still present our findings for (14) and (15) because the estimates for these can serve as useful benchmarks and also because the test (41) requires $\gamma_{j2} - \gamma_{j1}$ to be the same for all j's, which is not necessary for (14) and (15).

Following (14), the CPRO α-columns in Table 3 are obtained by transforming the μ-columns:

$$\alpha_{21} = \mu_{21} \cdot 1.937, \quad \alpha_{23} = \mu_{23} \cdot 1.940, \quad \alpha_{24} = \mu_{24} \cdot 1.439,$$
$$\alpha_{41} = \mu_{41} \cdot 1.346, \quad \alpha_{42} = \mu_{42} \cdot 0.695, \quad \alpha_{43} = \mu_{43} \cdot 1.349.$$

Since α_{jm} differs from μ_{jm} only in scale, the t-values for the μ-columns apply to the α-columns as well.

Based on the CPRO α-column and the estimates for the exogenous variables in the μ-column, the CPRO y_2-SF equation (male market labour) can be interpreted as follows. One hour increase in male farm-labour supply decreases the male market-labour supply by 1.154 hours; this estimate is statistically significant. The female's labour-supply decisions make little difference to the male, and their effects are not statistically significant. As usual, age has a positive sign and age2 has a negative sign, yielding an inverted quadratic shape peaking at age 41. Among the other regressors, only ed1c seems relatively important in terms of statistical significance; it has the expected positive effect on male market-labour supply.

Similarly, among the ODR endogenous variables in the CPRO y_4-SF (female market labour) equation, only y_3 (female farm labour) has a statistically significant effect. The effect is negative as in the case of males, but is much stronger (−2.345). A possible reason for this could be that farm labour supply and home production are complementary for females, so an increase in the former implies an increase in the latter as well, hence market labour supply is reduced further.

Table 3. Constant-threshold (CPRO) and varying-threshold probit (VPRO).

	CPRO				VPRO			
	y_2 SF (male market)		y_4 (female market)		y_2 SF (male market)		y_4 (female market)	
	μ (t-value)	α	μ (t-value)	α	μ (t-value)	α	μ (t-value)	α
y_1	−0.596 (−4.61)	−1.154	−0.147 (−0.27)	−0.198	−0.515 (−3.30)	−0.858	−0.370 (−0.46)	−0.567
y_2			−0.808 (−0.82)	−0.562			−1.199 (−0.65)	−0.835
y_3	−0.090 (−0.21)	−0.175	−1.738 (−2.83)	−2.345	0.230 (0.49)	0.363	−2.069 (−1.38)	−3.745
y_4	0.272 (0.77)	0.391			0.657 (2.19)	0.784		
1	−0.560 (−0.49)		−4.340 (−2.23)		1.799 (1.16)		−6.405 (−1.71)	
age	0.822 (1.62)		2.314 (1.84)		−0.441 (−0.67)		2.990 (1.66)	
age2	−0.101 (−1.87)		−0.257 (−1.78)		0.033 (0.50)		−0.314 (−1.62)	
ed1h	0.089 (1.07)				0.092 (0.93)			
ed1c	0.223 (1.53)				0.353 (2.04)			
ed2h			0.209 (0.82)				0.386 (0.69)	
ed2c			0.272 (0.89)				0.515 (0.72)	
nkid	−0.015 (−0.60)		−0.067 (−1.37)		−0.015 (−0.57)		0.012 (0.17)	
nado	−0.006 (−0.15)		0.011 (0.19)		0.035 (0.70)		−0.075 (−0.71)	
nadu	−0.026 (−0.69)		−0.120 (−1.68)		0.029 (0.78)		−0.085 (−0.93)	
is1	−0.103 (−0.95)				−0.150 (−1.25)			
aa1	0.090 (1.12)				−0.008 (−0.084)			
is2			0.307 (2.34)				0.323 (1.26)	
aa2			0.087 (0.67)				0.345 (0.86)	
	RF log-like: −1222.0, −1211.2, −1515.6, −1489.0				RF log-like: −1189.3, −1148.3, −1473.8, −1438.9			

As in the male equation, age and age2 show the same bell-shaped pattern, peaking at age 45. The number of children and the number of adults seem to have negative effects; the reason may be that they increase the demand for female time in home production. Israeli-born females are more active in market labour supply. With (15), we can identify the ratio of the coefficients of any exogenous regressor appearing in both y_2-SF and y_4-SF. For instance, for nadu, the ratio is $\frac{-0.026}{-0.120} \cdot \frac{0.433}{0.301} = 0.312$: The effect of nadu on male market labour is only 31.2% of that on female market labour.

The right half of Table 3 presents the VPRO estimates. We use the same regressors in both the regression function and two thresholds, since there is no good reason to rule out any regressor in the regression function from the varying thresholds; this results in $17 \times 3 = 51$ parameters. In this case, the endogenous SF parameters are identified, but not the exogenous SF parameters, according to the identification result in Section 3.2.

Using the following test statistic values for the y_2 and y_4 RFs, the null hypothesis of constant thresholds is easily rejected (p-values are all smaller than 0.00):

Likelihood ratio test (dof = 32)	: 125.8, 100.3
Score test (dof = 32)	: 83.45, 85.71
ODR-collapsing Wald test (dof = 16)	: 210.9, 63.2.

As for the threshold-difference constancy test for (42) with VPRO, we ran into a problem implementing the idea: some ratios obtained with MDE took negative values, because there is nothing built-in in estimation to make sure that the signs of $(\theta_{j3} - \theta_{j2})^{(c)}/\sigma_j$ and $(\theta_{m3} - \theta_{m2})^{(c)}/\sigma_m$ agree for all c and (j, m). This should be taken as an evidence rejecting the null hypothesis of the threshold-difference constancy. Nevertheless, as in the constant threshold case, we will still proceed imposing the threshold-difference constancy.

Since the actual thresholds $z_i' \, (\theta_{jr} - \theta_{j1})/\sigma_j$ are assured to be ordered in estimation ($0 < z_i' \, (\theta_{j2} - \theta_{j1}) < z_i' \, (\theta_{j3} - \theta_{j1}), \ldots$), the ratios that are always positive are

$$\frac{z_i'(\theta_{j3} - \theta_{j2})/\sigma_j}{z_i'(\theta_{m3} - \theta_{m2})/\sigma_m}, \quad i = 1, \ldots, N.$$

Now, there are N-many estimates for the single ratio σ_m/σ_j. The equally weighting MDE in this case is the simple average of the N-many estimates. Implementing this, we get

$$\alpha_{21} = \mu_{21} \cdot 1.666, \quad \alpha_{23} = \mu_{23} \cdot 1.579, \quad \alpha_{24} = \mu_{24} \cdot 1.194,$$
$$\alpha_{41} = \mu_{41} \cdot 1.532, \quad \alpha_{42} = \mu_{42} \cdot 0.696, \quad \alpha_{43} = \mu_{43} \cdot 1.810.$$

The VPRO α-columns were obtained using these.

The ability to interpret the magnitude of the SF estimates under (13) or (28) makes inference from the data analysis easier and richer than otherwise. Even when (13) and (28) do not hold exactly, the SF estimates obtained under them still provide 'reference points', much as a maximum-likelihood estimator often does in practice, even when the assumptions for the maximum-likelihood estimator do not hold exactly.

6. CONCLUSIONS

In this paper, we addressed the identification issue for simultaneous equations with ordered discrete response (ODR) variables, for varying- as well as constant-threshold specifications. One by-product of this analysis was a much simpler presentation of the usual order and rank identification conditions for simultaneous equations. Furthermore, one of the interesting findings was that by taking advantage of a simple cross-equation restriction, it is possible to fully identify the structural-form parameters of the endogenous regressors. Tests for this cross-equation restriction as well as for threshold constancy were proposed. Our identification results and proposed tests were applied in an empirical example of joint labour-supply decisions in farm households, and the magnitudes of the endogenous structural form parameters were interpreted accordingly, despite the unobserved scale factors in ODR.

ACKNOWLEDGEMENTS

The authors are grateful to Pravin Trivedi and an anonymous referee for their helpful and constructive comments.

REFERENCES

Bartholomew, D. J. and M. Knott (1999). *Latent Variable Models and Factor Analysis*. 2nd ed., Vol. 7. Arnold: Kendall's Library of Statistics.

Carlson, J. A. and W. C. Dunkelberg (1989). Market perceptions and inventory price employment plans. *Review of Economics and Statistics 71*, 318–24.

Kimhi, A. (2004). Family composition and off-farm participation decisions in Israeli farm households. *American Journal of Agricultural Economics 86*, 502–12.

Kimhi, A. and M. J. Lee (1996). Joint farm and off-farm work decisions of farm couples: estimating structural simultaneous equations with ordered categorical dependent variables. *American Journal of Agricultural Economics 78*, 687–98.

Lee, M. J. (1995). A semiparametric estimation of simultaneous equations with limited dependent variables. *Journal of Applied Econometrics 10*, 187–200.

Lee, M. J. (2002). *Panel Data Econometrics: Methods-of-Moments and Limited Dependent Variables*. San Diego: Academic Press.

Machin, S. J. and M. B. Stewart (1990). Unions and the financial performance of British private sector. *Journal of Applied Econometrics 5*, 327–50.

McIntosh, J., F. Schiantarelli and W. Low (2000). A qualitative response analysis of UK firms' employment. *Journal of Applied Econometrics 15*, 367–99.

Melenberg, B. and A. Van Soest (1996). Measuring the cost of children: parametric and semiparametric estimators. *Statistica Neerlandica 50*, 171–92.

Nadeu, R., R. G. Niemi and T. Amato (1995). Emotions, issue importance, and political learning. *American Journal of Political Science 39*, 558–74.

Pudney, S. and M. Shields (2000). Gender, race, pay and promotion in the British nursing profession: estimation of a generalized ordered probit model. *Journal of Applied Econometrics 15*, 367–99.

Terza, J. V. (1985). Ordinal probit: a generalization. *Communication in Statistics: Theory and Method 14*, 1–11.

Wansbeek, T. and E. Meijer (2000). *Measurement Error and Latent Variables in Econometrics*. Amsterdam: North-Holland.

Watts, M. and G. J. Lynch (1989). The principles courses revisited. *American Economic Review 79*, 236–41.

Windmeijer, F. A. G. and J. M. C. Santos-Silva (1997). Endogeneity in count data models: an application to demand for health care. *Journal of Applied Econometrics 12*, 281–94.

Winter-Ebmer, R. and J. Zweimuller (1997). Unequal assignment and unequal promotion in job ladders. *Journal of Labour Economics 15*, 43–71.